用数据驱动选题策划
——试论图书编辑如何运用大数据背景下的信息资源开展选题工作

刘万年
（北京体育大学出版社，北京　100000）

摘　要：本文设计一种基于大数据的选题调研方法，辅助图书编辑从研究大数据背景下的信息资源开展选题策划工作。首先分析图书编辑领域需求，了解出版行业选题策划背景；其次基于大数据采集图书信息，保证图书数据的真实性；接着展开数据驱动选题策划的可行性分析，从市场需求与供给两个层面优化图书选题方向，提升图书选题的受欢迎程度，从而实现由数据驱动的选题策划模式来辅助编辑的选题工作与决策。最后采用对比实验的方式证明了这一方法具有可研究和推广价值。

关键词：数据驱动；选题策划；图书编辑；大数据背景；信息资源　　**中图分类号：**G256.1　　**文献标识码：**A
文章编号：1671-0134（2022）05-120-03　　DOI：10.19483/j.cnki.11-4653/n.2022.05.037
本文著录格式：刘万年 . 用数据驱动选题策划——试论图书编辑如何运用大数据背景下的信息资源开展选题工作 [J]. 中国传媒科技，2022（05）：120-122.

导语

全球经济正在迅猛发展，人们进入了大数据时代。面对庞大的图书市场，图书可选信息越发繁杂，图书编辑如何从其中筛选出符合市场需求的信息是极为重要的工作。[1] 因为一旦图书编辑没有了解到市场的真实需求，就会导致所策划的图书无法得到市场认可，从而对销量产生巨大影响，直接威胁着图书企业的生存发展。图书编辑作为一个肩负着传播知识的历史使命的传统行业从业者，在此大环境下同样面临着大数据时代带来的新挑战。[2] 当前，我国出版业面临着图书消费需求增长变缓，图书成本提高、库存加重等问题，造成图书行业面临很大的发展困境。因此图书行业迫切需要深化改革，其中图书选题的进一步优化是改革重点之一，选题策划是图书编辑进行决策的关键步骤，一本图书的成功发行，离不开准确有效的选题策划。

图书编辑可以根据出版社的选题方向、市场需求、读者体验等大数据分析结果进行具体分析规划。在图书的多样化、个性化、社会化的特征基础上，准确了解大众的阅读倾向和阅读需求，分析出图书消费市场的风向动态，及时搜集整理图书市场的热点信息，最大限度地判断图书市场的趋势，有效掌握公众反馈信息，为选题决策提供依据。

1. 基于大数据的选题调研案例设计

1.1　分析图书编辑在选题策划中的数据需求

图书市场的价格趋势，读者购买量和阅读习惯，图书各渠道销售情况的收集、处理和分析均来自图书市场的反馈，这其中蕴含着很多机遇，进而形成新的图书选题策划。[3] 图书编辑要对市场趋势、消费者反馈、竞品图书、推广方式、盈利能力等信息进行了解，需要保证图书信息的准确。在当前的图书市场中，图书信息在各大购物网站上分布广泛，图书销售的数据量大、增长速度快，数据获取随之增加。要想迅速识别图书市场的有用信息，利用大数据技术开展图书市场需求的分析就可以高效达成，且能进一步保证数据的精准度。基于此，本文使用大数据背景下的 Hive 数据管理技术，对图书数据进行管理，完整的模块信息如下图 1 所示。

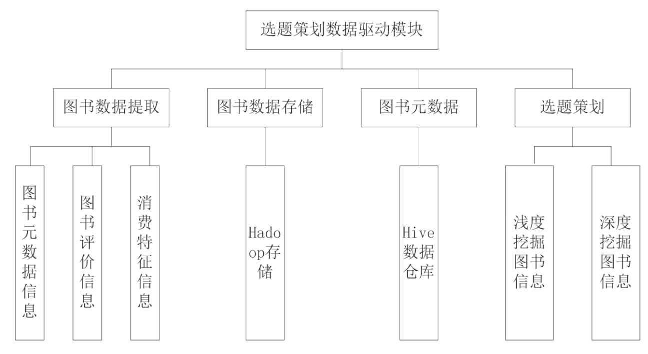


图 1 基于大数据的图书选题策划信息模块

如图 1 所示，在大数据的应用下，图书选题策划数据驱动模块可以分为图书数据提取模块、图书数据存储模块、图书元数据模块、选题策划模块四部分，每个模块的特征向量均会存在不同的权重。由于图书购买用户不同，根据用户信息可以分析出用户的需求书籍类型，在图书买卖过程中，会出现不同浏览行为，通过图书翻阅次数、图书购买次数、图书评价等信息，可以指导图书的选题策划方向。[4] 此外，本文设计的图书选题策划方法，会根据图书近期销售进行预测，如果对用户的预测成功，则说明策划方案是成功的；如果失败，则需要进一步分析图书购买需求，完善选题策划的多样性与丰

富性。在使用大数据技术过程中，需要对用户的图书购买特征进行分析，去除从众因素，保证图书数据的真实性。根据大数据分析结果对图书市场需求进行精准把控。

1.2 基于大数据采集选题信息

在大数据时代，人们的需求开始透明化，通过大数据的使用，将用户对图书的购买行为进行分析，了解图书市场的动态变化。在图书编辑领域，选题工作不仅要充分利用大数据，还要对图书购买需求进行预测，扩展数据所反映的图书价值。

大数据的影响越来越广泛，已经逐渐渗透到人们生活的方方面面。[5] 大数据正在改变着人们的生活方式，其数据信息处理的结果，有可能比业内资深人士预测得更加准确。传统的选题信息采集工作，局限于策划者的信息收集，高度依赖于优秀编辑的经验判断。有些时候，图书编辑人员会出现错失良机的现象，导致图书信息采集不准确，对同期的图书销售造成影响。而本文基于大数据对图书选题信息进行采集，通过数据的全面收集和分析，让数据说话，准确地预测图书销量信息与受欢迎程度。[1]

在此预测环境下，可以增强图书信息的选题准确性，并满足读者需求。大数据的采集工作，主要通过对图书市场的精准分析，完整地把控图书市场需求，同时反向运作也同样适用，如通过对不同年龄段的喜好，进行书籍的推送。以青少年为例，可以推送一些具有丰富科学知识的小说，不仅可以增加阅读兴趣，还可以学习到相关知识。对中老年人来说，多推送一些她们年轻时代的故事，可以使其回忆起年轻的生活，并对未来的道路充满希望。我国目前有 500 多家出版社以及相当数量的民营图书策划公司，图书编辑队伍庞大，但是图书选题的主题相对欧美国家要少，每年畅销书数量与整体行业销售量不成正比。本文认为，主要是因为图书编辑没有对市场的需求形成正确的理解把握，获取到的信息不能够支撑编辑团队的真正成长。大数据时代的到来，为图书编辑指明了方向，那就是利用数据支撑，收集有效准确的选题信息，切实策划符合市场需求的图书产品，并提高选题质量，促进图书市场的繁荣与发展。

1.3 进行数据驱动选题策划的可行性分析

从市场层面分析，教育类图书的编辑主题与内容框架，可以通过筛选教育数据，将其进行供应需求的分析。由于我国对教育行业的重视程度，图书市场中教育类图书所占比重较大，在庞大的教育图书中，图书编辑必须提前解决选题内容框架，减少潜在的竞争。因此，对教育类的图书选题策划过程中，符合市场要求与竞争的有效资源信息方面至关重要。从供给层面分析，需要图书编辑参与图书数据库的搭建。[6] 并利用专业化的指导和时效性的数据，对教育类图书的选题进行调研，并利用优秀编辑相关教育图书方面的编辑经验，对选题内容进行初步规划，再利用大数据深度挖掘图书信息，真正意义上满足图书市场的发展需求。此外，大数据的应用经过多年积累，可以称其为大数据智能平台，通过大数据对用户的购买行为来获取数据情报，得到图书的有效价值信息。使用大数据的后台技术，对相关用户的图书购买行为进行记录，收集用户的搜索关键字，使图书选题策划在行业领域内拥有领先优势。任何产业的发展，都会结合最新的尖端技术，图书行业也不例外，通过大数据的融合，可以将图书数据精准地把控，得出准确的数据基础，保证图书选题策划的精准性，对图书行业的发展创造潜力条件。综上所述，本文设计的图书选题策划方法，通过大数据的介入，为图书编辑提供一个接地气的、强有力的、具有真实市场指导意义的辅助工具。

1.4 实现数据驱动的选题策划模式

为了实现数据驱动的选题策划模式，本文设计了在数据挖掘下的数据库，如下表 1 所示。

表 1 核心数据库

数据类型	核心数据
PK	bookID_scoreStar dateNum_location pricecurrent
FK	bookname_isbncode publisher_pageNum versionnumber price_seriensname
FK1	reviewed_bookDate sorceStar userLevel_location praiseGood
Bookid	type_scoreDate location level

如表 1 所示，为本文设计的核心数据库，根据此数据库的设计，可以提取图书相关信息，通过图书 ID，建立图书链接，从而得到相关数据的属性信息。

功能数据库设计如下表 2 所示。

表 2 功能数据库

数据类型	功能
Function	buildBookUrl
Description	拼装图书高级搜索的 URL
Param	isbnCode，图书的 ISBN 号
Return:	图书高级搜索中对应 ISBN 号的 URL

表 2 所示为功能数据库的设计，通过搜索图书 URL，得出相关图书的信息数据。

数据存储库的设计如下表 3 所示。

表 3 数据存储库

数据类型	存储位置
Studo vi	/etc/sysconfig/netwok
Vi	/etc/hosts
Studo	service iptables stop
Ssh-keygen-t rsa	~/.ssh/id_rsa.pub
Cp	~/.ssh/authorized_keys
Hadoop	namenode-format start-all.sh

通过此数据库的设计，可以实现选题策划的信息读取、更新、删除，以及增加等操作，保证图书数据的丰富性。

2. 实验

为了验证本文设计方法是否具有实效性，在此对其展开实验，实验的过程及结果如下。

2.1　实验准备

由于图书市场环境较为复杂，本文利用分布式将图书信息集群分为 5 个节点，使用 Master 与 CentOS 为主要操作系统，以 Linux 为主要服务器，CPU 主频保持在 3.5GHz，信息数据存储为 8G，存储空间为 1T，由此得出的集群参数如下表 4 所示。

表 4　图书集群参数表

序号	节点名称	集群 IP	ID	数据存储位置
1	cloud1	10.5.110.242	hadoopusr	/opt/hadoop-1.2.1
2	cloud2	10.5.110.243	hadoopusr	/opt/hadoop-1.2.1
3	cloud3	10.5.110.244	hadoopusr	/opt/hadoop-1.2.1
4	cloud4	10.5.110.245	hadoopusr	/opt/hadoop-1.2.1
5	cloud5	10.5.110.247	hadoopusr	/opt/hadoop-1.2.1

如表 4 所示，在此参数下，对图书数据进行可行性评估。为了提高实验的精准度，本文从图书数据量的 100 万行至 1000 万行之间的图书进行数据分析，在大数据背景下，得出图书信息集群的 5 个节点图书选题受欢迎程度如下图 2 所示。

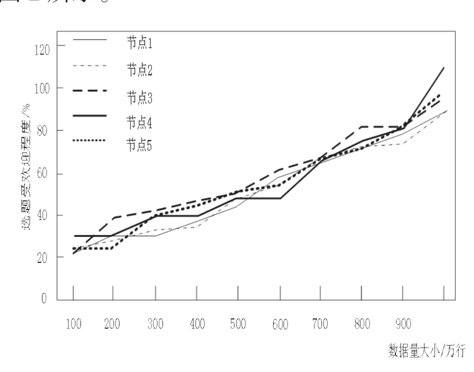


图 2 图书信息集群的 5 个节点图书选题受欢迎程度

如图 2 所示，在 5 个节点中，随着数据量的增加，图书选题受欢迎程度也随之增加，可以满足选题需求。

2.2　实验结果

在以上环境下，对传统选题策划方法图书选题受欢迎程度，与本文设计选题策划方法图书选题受欢迎程度做对比，结果如表 5 所示，两种方法图书选题受欢迎程度均会随着数据量的增加而增加。传统选题策划方法图书选题受欢迎程度始终在 60% 以下，受欢迎程度较低，无法适应大数据环境；本文设计的选题策划方法图书选题受欢迎程度在数据量为 1000 时，理论上受欢迎程度达到了 100%，可以适应大数据环境，保证选题的精确性。符合本文研究目的。

表 5　实验结果

数据量	传统选题策划方法图书选题受欢迎程度	本文设计选题策划方法图书选题受欢迎程度
100	8.90%	35.62%
200	12.41%	40.25%
300	15.48%	46.31%
400	19.25%	52.15%
500	21.34%	62.28%
600	24.36%	74.26%
700	31.62%	88.92%
800	35.46%	92.98%
900	40.21%	99.82%
1000	56.23%	100%

结语

近年来，在大数据的背景下，各行各业得到了空前的发展，为我国经济进步创造了条件。图书编辑作为人们学习知识的铺路者，对推动人类精神文明的建设起到了相当重要的作用。传统图书选题策划较为主观化，其选题受欢迎程度平均水平较低，目前已经不再适应图书传播需求，无法巩固社会经济效益。基于此，本文以大数据为前提，设计选题策划方法，摒弃传统方法的缺点，通过与大数据的融合，对图书选题进行可行性分析，科学指导市场图书选题策划方向，提高选题受欢迎程度，以期为今后图书行业的深化改革与建设添砖加瓦。媒

参考文献

[1] 姜海涛 . 浅析融合出版选题策划的多维意识——以《身边汉字》融合出版项目为例 [J]. 科技与出版，2020（2）：72-77.

[2] 崔美花 . 新媒体环境下广播电台节目编辑与策划创新分析 [J]. 中国传媒科技，2019（9）：91-93.

[3] 魏晓平 . 图书策划编辑中“互联网 +”思维的运用 [J]. 中国传媒科技 2021（12）：75-77.

[4] 毋婷娴 . 论科技类图书编辑如何选题策划 [J]. 新闻研究导刊，2021（7）：201-202.

[5] 张娓莹 . 浅析信息化环境下图书编辑做好选题策划的策略 [J]. 传播力研究，2020（3）：124-125.

[6] 毛雪 . 数字时代图书编辑如何创新 [J]. 新媒体研究，2016（4）：127-128.

作者简介： 刘万年（1975-），男，河北张家口，中级编辑，研究方向：编辑学。

（责任编辑：胡杨）